Unit II

11. Attribute Aggregation:

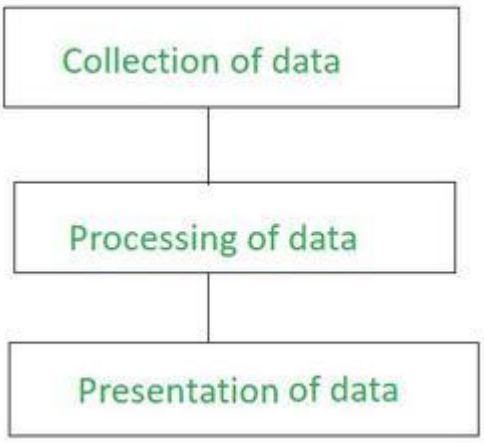
Attribute aggregation, also known as feature aggregation or variable aggregation, is a data preprocessing technique in data analytics that involves combining or summarizing multiple attributes or features into a single attribute. The purpose of attribute aggregation is to simplify the data representation and reduce the dimensionality of the dataset by consolidating related information into a more manageable and meaningful form.

In data analytics, attribute aggregation is commonly used for the following reasons:

1. Reducing Dimensionality: Large datasets with numerous attributes can be computationally expensive and may lead to overfitting in certain modeling algorithms. Attribute aggregation helps in reducing the number of variables, making the data more manageable and efficient for analysis.
2. Data Simplification: Aggregating attributes allows data analysts to focus on higher-level patterns and trends rather than individual details. It simplifies the data representation, making it easier to interpret and visualize.
3. Handling Sparse Data: In some cases, certain attributes may have a large number of missing or sparse data points. Attribute aggregation can be used to consolidate sparse attributes and provide a more comprehensive view of the data.
4. Improving Model Performance: In machine learning and statistical modeling, aggregating correlated attributes can lead to better model performance by removing multicollinearity (high correlation) among the predictors.

Common techniques for attribute aggregation include:

1. Summation: Aggregating numerical attributes by adding their values. For example, combining sales data for individual products into total sales for a specific category.
2. Averaging: Calculating the average of numerical attributes. This can be useful when dealing with repeated measurements or survey responses.
3. Counting: Aggregating categorical attributes by counting the occurrences of each category. For instance, counting the number of occurrences of specific keywords in text data.
4. Grouping: Grouping data based on certain criteria and summarizing the values within each group. Grouping can be used to create aggregated statistics, such as mean, median, or sum, for each group.
5. Max/Min: Taking the maximum or minimum value of numerical attributes within a group or over a specific period.
6. Feature Engineering: Creating new features by combining existing attributes using mathematical operations, ratios, or percentages.



How do data aggregators work?

Data aggregators work by combining atomic data from multiple sources, processing the data for new insights and presenting the aggregate data in a summary view. Furthermore, data aggregators usually provide the ability to track data lineage and can trace back to the underlying atomic data that was aggregated.

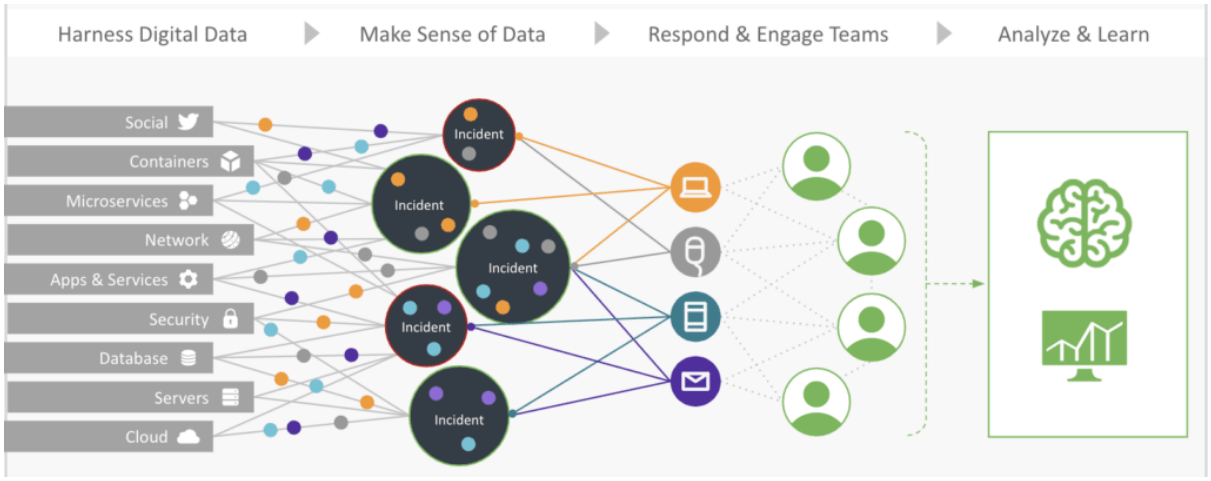
Collection. First, data aggregation tools may extract data from multiple sources, storing it in large databases as atomic data. The data may be extracted from internet of things (IoT) sources, such as the following:

- social media communications;
- news headlines;
- personal data and browsing history from IoT devices; and
- call centers, podcasts, etc. (through speech recognition).

Processing. Once the data is extracted, it is processed. The data aggregator will identify the atomic data that is to be aggregated. The data aggregator may apply predictive analytics, artificial intelligence (AI) or machine learning algorithms to the collected data for new insights. The aggregator then applies the specified statistical functions to aggregate the data.

Presentation. Users can present the aggregated data in a summarized format that itself provides new data. The statistical results are comprehensive and high quality.

Data aggregation may be performed manually or through the use of data aggregators. However, data aggregation is often performed on a large-scale basis, which makes manual aggregation less feasible. Furthermore, manual aggregation risks accidental omission of crucial data sources and patterns.


Harness Digital Data
Make Sense of Data
Respond & Engage Teams
Analyze & Learn
Social
Containers
Microservices
Network
Apps & Services
Security
Database
Servers
Cloud
Incident
Incident
Incident
Incident
Incident